DOCUMENT RESUME

ED 384 222 FL 023 051

AUTHOR	Grant, Leslie
TITLE	Testing Bilingual Teachers' Language Proficiency: The Case of Arizona.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	Jun 95
NOTE	27p.
PUB TYPE	Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Bilingual Education; Educational Policy; Factor Analysis; Interrater Reliability; *Language Proficiency; *Language Tests; Limited English Speaking; Questionnaires; Second Language Learning; Spanish; Tables (Data); *Teacher Certification; Teacher Qualifications; Testing; Test Validity
IDENTIFIERS	*Arizona

ABSTRACT

Many states currently offer bilingual certification or endorsement, encouraging both practicing teachers and prospective teachers to complete their requirements necessary to add this certification on to their regular teaching license. Although these requirements routinely include courses in bilingual education, the "second" language proficiency of the teacher is only sometimes addressed. The purpose of this paper is to discuss proficiency testing of bilingual teachers across the United States. After discussing testing of bilingual teachers in general, a specific measure used in Arizona will be described and evaluated. A sample questionnaire is appended. (Contains 19 references.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Testing Bilingual Teachers' Language Proficiency: The Case of Arizona

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Leslie Grant

Educational Testing Service

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Leslie Grant
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 384 222

Abstract

Many states currently offer bilingual certification or endorsement, encouraging both practicing teachers and prospective teachers to complete the requirements necessary to add this certification on to their regular teaching license. Although these requirements routinely include courses in bilingual education, the *second* language proficiency of the teacher is only sometimes addressed. The purpose of this paper is to discuss proficiency testing of bilingual teachers across the United States. After discussing testing of bilingual teachers in general, a specific measure used in Arizona will be described and evaluated.

Introduction

The measurement of bilingual teachers' language proficiency, a specific kind of language ability, is a concern for many areas in the United States. The number of limited English proficiency (LEP) students continues to rise as does the need for bilingual teachers, particularly in those areas where bilingual education is an option for meeting the needs of the students. Of the approximately 45 million school age students, about 6.3 million have been designated to be LEP; the most common native language of this group is Spanish (Waggoner, 1992). A forum held by the Office of Bilingual Education and Minority Language Affairs (OBEMLA) addressed staffing needs for this decade. The participants noted that the "rapid increases in the number of LEP students is compounding the existing problem of bilingual/ESL staff shortages" (Esquivel, Ovard, Wooten and Wilkisin, 1990, p.3).

A partial response to the need for bilingual teachers has come from the institutions of higher education. Many colleges and universities have incorporated bilingual education option as part of their teacher training programs. In addition to the classes required of these potential

FL023251

BEST COPY AVAILABLE

bilingual teachers, it is of course necessary that potential teachers meet certain language proficiency standards.

Review of Bilingual Teacher Testing in the U.S.

How then is the proficiency of bilingual teachers being evaluated across the United States? The answer is that there is great disparity in how teachers are assessed. Presently, out of the 50 states in the U.S., 28 offer either certification or an endorsement in bilingual education. Table 1 reports these states, the type of certification or endorsement that is offered, and the type of test required for certification or endorsement. As is indicated in the table, of these states, 11 (Alaska, Connecticut, Indiana, Kansas, Louisiana, Minnesota, North Dakota, Ohio, Oklahoma, Vermont, and Washington) solely require teachers to complete coursework to receive their certification or endorsement. That is, as candidates complete their programs, they are eligible for certification from the state; no test of language proficiency is required.

Seven states (Colorado, Delaware, Idaho, Maine, New Hampshire, New York, and Wisconsin) leave the decision of how to assess language skills up to the teacher preparation institutions. Institutions may require demonstration of language proficiency by using a multiple choice test, a written essay, and/or an oral proficiency measure. Often times, the completion of a number of target language credit hours serves to fulfill the language requirement. Thus, completion of this requirement may result in essentially no assessment of language abilities (outside of the language classroom). Sometimes, however, the institution may require that the candidates take a language proficiency test or tests to demonstrate competence in the target language.

Of the remaining ten states, four use a single test to assess language. Michigan currently requires teachers seeking certification to pass a multiple choice test, consisting of questions (in English) on the topic of bilingual education and theories of second language acquisition (Dorothy

Table 1

Certification/Endorsement and Testing Procedures for Bilingual Teachers in the U.S.

State	Certif/Endor	Test Required
Alaska	E	None
Arizona	E	Arizona Classroom Teacher Spanish Proficiency Exam
California	C	Bilingual Crosscultural Language and Academic Development Examination (BCLAD)
Colorado	E	Institute of Higher Education's choice
Connecticut	C	None
Delaware	C	Institute of Higher Education's choice
District of Columbia	C	Language Proficiency Interview
Idaho	C	Institute of Higher Education's choice
Illinois	C	Language Proficiency Interview; Written Test
Indiana	E	None
Kansas	E	None
Louisiana	E	None
Maine	E	Institute of Higher Education's choice
Massachusetts	E	English Exam, Target Language Exam, Culture Exam
Michigan	C	Written Multiple Choice Exam
Minnesota	C	None
New Hampshire	E	Institute of Higher Education's choice
New Jersey	E	Target Language Assessment
New Mexico	C	New Mexico Four Skills Exam
New York	C	Teacher Preparation Institute's Choice
North Dakota	C	None
Ohio	E	None
Oklahoma	E	None
Rhode Island	E	Language Proficiency Interview
Texas	C/E	ExCET (Bilingual Ed. Content Area Test); Texas Oral Proficiency Test
Vermont	E	None
Washington	E	None
Wisconsin	C	Institute of Higher Education's choice

Note. Information from McFerren, Valadez, Crandall, Paloma, & Patino-Gregoire, 1988; National Clearninghouse for Bilingual Education, 1991.

Van Loy, personal communication, February, 1995). The single measures used by three states, District of Columbia, Rhode Island, and New Jersey, is a language proficiency interview. This type of testing most often follows the format of the American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview (American Council on the Teaching of Foreign Languages, 1986) and reports scores using the ACTFL scale (or at times, the Foreign Service Institute scale). In utilizing solely this type of testing procedure, these states indicate that a teacher's ability to speak in the target language is the most important consideration.

Finally, the remaining six states require multiple measures for bilingual teacher certification and/or endorsement. Illinois requires an oral interview, a multiple choice reading comprehension exam, a written essay test, and a cloze test (Jolene Reddy, personal communication, February, 1995). Massachusetts reports having an exam in English, an exam in the target language, and a culture exam (National Clearinghouse for Bilingual Education, 1991) while the states of Arizona, California, New Mexico and Texas have developed their own exams. The fact that these states have multiple measures suggests that they feel that a bilingual teacher needs to demonstrate more than oral proficiency in order to be endorsed or certified.

The exams used in the last states mentioned above reflect in-state efforts alone, or the collaboration of state personnel and consultants, for exam development. Two of these states, New Mexico and Texas, require that examinees engage in measures that test their language skills; efforts to keep the tests general were made. In other words, these tests address the language ability of the examinees, but the tasks are not as specific as others that are found in performance-based tests, for example (Wesche, 1992). The tests required by Arizona and California, in contrast, have determined tasks that are representative of "teacher tasks," yet the approaches for test development differed. In the case of Arizona, a needs analysis which addressed the language needs of bilingual teachers was conducted (Barkin-Riegelhaupt, 1985). In California, a survey was used and teachers indicated which among a list of teacher tasks were most relevant for a certification test (Cooperative Personnel Services, 1992). Thus, the tests in Arizona and

California targeted a specific type of language proficiency in the design of their tests -- the proficiency needed by bilingual teachers for teaching.

Arizona's Spanish proficiency test

In Arizona, an estimated 60,000 of the school age children are LEP (which is about 9% of the school population). School districts, by law, must provide these children designated as LEP with either a program in bilingual instruction or English as a second language (ESL). In addition, all personnel who work as bilingual classroom instructors must have a bilingual education endorsement (Arizona Department of Education, 1992). In order to receive such an endorsement, certain criteria must be met, one of which is passing a language proficiency test.

El Examen de Proficiencia en Español Pedagógico, or the Arizona Classroom Teacher Spanish Proficiency Exam (ACTSPE), was developed by the Arizona Department of Education in the mid-1980s in cooperation with the three universities which also serve as testing sites for the exam: Arizona State University (ASU), Northern Arizona University (NAU), and University of Arizona (UA). This test is a criterion-referenced, partially performance-based test. The main objective of the ACTSPE is the evaluation of the examinee's ability to use Spanish in the bilingual classroom. As a result, the test items are based on the Spanish that is used in the daily activities of the teacher; realistic tasks such as correcting students' writing, translating a letter home to parents, reading aloud to children, and teaching a lesson are used to measure the examinee's level of teaching proficiency in Spanish.

This unique mix of items is the result of a needs analysis that was conducted in the early 1980s (Barkin-Riegelhaupt, 1985). Representatives from various areas such as parents, children, administrators, and test designers provided information regarding the language needs of the bilingual teacher. This information, in combination with bilingual teacher observations, was utilized to outline the tasks that bilingual teachers need to be able to perform. For example, teaching a lesson, translating a letter to parents, and correcting students' writing were determined to be tasks on the ACTSPE by means of this process.

What do the tasks on the ACTSPE measure with respect to language ability? At first glance, it appears that the test taps several competencies found in a model such as Bachman's Communicative Language Ability (1990). Some of the "performance-based" sections on the test require the examinee to demonstrate sociolinguistic competence through oral language use (e.g., as in the telephone conversation with a parent--Section 1), while other sections tap the grammatical competence of the examinee (e.g., Section 8--Correcting students' writing). Upon closer examination of the test, however, it is evident that many aspects of language competence come into play in nearly every test section, making the simple interpretation of the phone conversation a test of a prospective teachers' sociolinguistic competence inadequate. The phone conversation also is graded for grammatical accuracy, for example, while Section 8--Correcting students' writing--draws on textual competence. In this section the examinee must process the student's writing sample in context so that underlined words and phrases can be understood and corrected if necessary. This overlap of contributing "competencies" can be found in many of the test sections due to the tasks required (as well as the scoring rubrics used). This makes it difficult to isolate what is being tested in terms of a model of language ability; indeed the test preceded the model just mentioned. Yet the question asked above is an important one. Efforts to better understand what is being measured by this test serve as research questions for this study.

Research questions

The high stakes nature of the ACTSPE results underscores the importance of evaluating this test. Thus, the following questions were posed to investigate the ACTSPE.

Reliability. The first two questions addressed the reliability of the exam.

1. Are the oral and written parts of the test internally consistent?
2. Do the raters score consistently?

Validity. The three questions below addressed evidence for the validity of the exam:

1. What did the examinees think of the test?
   (a) Did the examinees think that the test diagnosed linguistic abilities of a teacher in a bilingual setting?
   (b) Did the examinees think that the test was at an appropriate level of difficulty for testing bilingual teachers?
2. Do the test tasks match the kinds of tasks required of a teacher in a bilingual classroom?
3. Do the separate parts of the test provide different information?

Method

Subjects

For this descriptive study, the scores of tests administered at the three testing sites were investigated. A convenience sample of 349 tests was available for this study (127 from ASU, 180 from NAU, and 42 from UA). No individuals *per se* were part of the study; however, results from questionnaires provided background information on the majority of the individuals whose scores were used. Of those who responded to the questionnaires (n = 285), 220 (77.2%) were female, and 65 (22.8%) were male. In addition, of the 201 who indicated their profession, 157 (78.1%) were already teachers, seven (3.5%) were students, while the other 37 (18.4%) were professionals (librarian or aide).

Instrumentation

The main measure used in this study was the Arizona Classroom Teacher Spanish Proficiency Exam (ACTSPE) which is described below. In addition to this test, two questionnaires were used.

Test description. The ACTSPE consists of two main parts: one oral and one written. The oral part has both speaking and listening tasks while the written includes both reading and writing. The test is given in a language laboratory and is administered in about three hours. Scores are reported in terms of an oral score and a written score. A description of the parts and procedures for scoring follow.

Oral part: The oral/aural component comprises the majority of the test (38 items; 62%). It is subdivided into six sections that the examinee must complete; the first five sections of the exam are tape-recorded. These sections are described below.

a) Section 1--Oral communication with parents

In this section the examinee records parts of a telephone conversation that have been omitted between a teacher and a parent. Prompts are provided for the examinee in the test booklet. Scoring of this section, as in many sections, is based on a scale. In this case the score is based on vocabulary, grammar, appropriateness, and naturalness (three points each for a total of 12 points).

b) Section 2--Oral reading in Spanish

In this section a reading selection of approximately 100 words is provided, with title and author. The examinee is instructed to study the passage, and then read it out loud as if reading to students in the classroom using expression, fluency, intonation, and rhythm appropriate for the section. This section is worth 5 points.

c) Section 3--Oral presentation of an instructional activity

In this section a paragraph has been chosen from a teacher's guide from a math or science series. The examinee is to teach the lesson presented in the paragraph to students as if he

or she were in the classroom. Examinees are scored on comprehensibility (eight points), grammar (five points), and fluency (two points). A total of 15 points is possible in this section.

d) Section 4--Formulation of questions

In this section the examinee formulates questions based on a story taken from a student reading text. The story is grouped into five sections and the examinee comes up with one question per section geared toward a group of students that have read the selection. Examinees are awarded two points per question for total of 10 points.

e) Section 5--Technical vocabulary

This oral section involves translation from English to Spanish. Included are numeric phrases, educational terminology, dates and hours, common phrases, and familial relationship terms. There are 20 items at one half point for each (total possible is 10 points).

f) Section 9--Listening comprehension of students

This final oral section involves students speaking in Spanish in the classroom and the examinee must indicate his/her comprehension of what they are talking about by answering multiple choice questions. These responses are written. Each question for both sections is worth one point and the 10 questions then make a total possible score of 10 points. Here the answers are written in the test answer sheet.

Written part: The reading/writing component consists of 26 items (38% of the test). It is comprised of three sections and all answers are written on the test answer sheet. The sections include:

a) Section 6--Translation of a letter to parents
A letter, in English, relaying important information regarding future school activities is t ranslated into Spanish in this section. A total of 12 points is possible and examinees are rated on grammar, vocabulary, spelling, and style.

b) Section 7--Written summary and reaction to professional reading
This section deals with an examinee's ability to read technical material in his/her area, to summarize its principle points, and to write a reaction to the content. A total score of 14 points is possible for this section and examinees are rated on comprehension, writing, and style.

c) Section 8--Written corrections of students' writing
Using authentic material written by students in bilingual programs, the examinee has to read compositions and rewrite the underlined words. The words that have been underlined may or may not need corrections. The examinee is told to use the spelling formally taught in textbooks. Spelling and accents are the primary ways that this section is scored. Twenty-four examples of writing are given and each is worth one half point. The total possible is 12 points.

Scoring: The scoring is done by two raters who evaluate each test independently. As described above, the majority of the sections utilize analytical scales and the scores for each of the scales are added together to yield a section score. The three sections that do not use a scale are scored dichotomously. These sections include two oral sections (Section 5--Technical vocabulary and Section 9--Listening comprehension of students) and one of the written sections (Section 8--Correcting students' writing). (Actual section scoring criteria are described in full in Norfleet, 1994.) After scoring has been completed, the oral section scores are added together and reported as are the written section scores. Passing is 70% on both parts; a bilingual endorsement is not granted until both parts are passed.

ACTSPE questionnaire. After the test is given, each test site administers an independently developed background questionnaire to find out who the examinees are in terms of first language,

opinion of the test and/or perceived level of appropriateness of the exam. Sample questions selected from all three questionnaires are found in Appendix A.

Teacher expert questionnaire. In addition to the exams and the exam questionnaire, another questionnaire was developed for this study to investigate the content of the exam. An example of the teacher expert questionnaire is found in Appendix B.

Procedures

Exams and related materials were collected from the three Arizona testing sites. In addition, 10 bilingual teacher experts (experts being defined as teachers having at least three years' experience in the classroom) were contacted to take part in the study.

Analyses

To address the five research questions described above, the data were analyzed both quantitatively and qualitatively where appropriate. Distributions, correlations, exploratory factor analysis, and reliabilities were computed using SAS (SAS Institute, Inc., 1985).

Results

Table 2 reports the test statistics for the ACTSPE. Overall, the oral and written parts of the exam appear to have been relatively easy for the examinees. The oral part mean score of 49.25 represents a percentage score of about 79% on the oral part (there are 62 points possible). The written part mean score of 27.08 represents a percentage score of approximately 71% (there are 38 points possible) for this part. These scores are more or less representative of scores on certification tests where the majority of the examinees take the exam as they near completion of their coursework and/or the program.

Table 2
Test Statistics for the ACTSPE

Variable	N	M	SD	Min	Max	Skew-ness	Kurto-sis
Oral Part	291	49.25	9.55	14	62	-1.30	2.40
Written Part	311	27.08	6.27	9	38	- .47	- .19

Reliability

Internal consistency. The oral and written parts were investigated for internal consistency using coefficient alpha (Bachman, 1990). Because of the nature of this test (oral and written parts) which can be administered, scored, and are reported separately, it was determined to treat the two parts as if they were, in effect, two different tests. The resulting reliability coefficients for the oral and written parts are found in Table 3.

Table 3
Coefficient Alpha for Oral and Written Parts of ACTSPE

Part	Reliability Coefficient
Oral	.87
Written	.88

Both the oral and written parts had high reliability coefficients, .87 and .88, respectively. These results suggest that the main parts of the test are consistently measuring the examinee's oral abilities on the oral part and writing abilities on the written part.

Interrater reliability. The final area of interest with respect to reliability was the rater consistency. Pairs of exam scores were only available at one site, NAU, so this analysis reflects those raters. In Table 4, the results from the correlation indicated that the raters were scoring consistently on the oral part of the exam. The criterion originally set for this certification test was .85. Rater pair 3 and 4, however, were slightly below that minimum.

Table 4
Pearson Product-Moment Correlations among Raters at NAU for Oral Part

	Rater #1	Rater #2	Rater #3
Rater #2	.93* (n = 26)		
Rater #3	.94* (n = 23)	.90* (n = 28)	
Rate #4	--	--	.84* (n = 42)

*$p \leq .05$

Correlations for the written part of the exam are found in Table 5. Again, the overall scoring is consistent, based on the .85 criterion, with the exception of the same rater pair, pair 3 and 4. These results suggest rater training is needed periodically, particularly for rater pair 3 and

4, to ensure that rater levels of agreement are meeting the standards set. This is especially important in a test of this nature, where many of the sections are scored subjectively.

Table 5
Pearson Product-Moment Correlations among Raters at NAU for Written Part

	Rater #1	Rater #2	Rater #3
Rater #2	.95* (n = 38)		
Rater #3	.95* (n = 23)	.88* (n = 36)	
Rate #4	--	--	.67* (n = 42)

*$p \leq .05$

Validity

Face validity. This research question addressed face validity. Two items from the exam questionnaire provided evidence for the face validity of the ACTSPE. For the first question (Did the examinees think that the test diagnosed linguistic abilities of a teacher in a bilingual setting?), approximately 81 examinees responded. The vast majority (n = 78, 96.3%) indicated that they felt that the ACTSPE was either very useful or moderately useful for diagnosing the linguistic abilities of a teacher in a bilingual setting. Very few (n = 3, 3.7%) thought that the exam was not useful at all.

The second question on the test questionnaire asked examinees if they felt that the test was at an appropriate level of difficulty for testing bilingual teachers. Of the 91 respondents, 77 (84.6%) felt the exam was appropriate. Ten respondents (11.0%) felt that the test was too difficult while 4 (4.4%) felt it was too easy.

Content validity. For this analysis, the test tasks were compared to the types of tasks that bilingual teachers typically experience in their classrooms. Ten bilingual teachers in Arizona were asked about the test content, section by section, to see if it represented the activities they face as bilingual teachers. Nearly all of the sections were perceived to be relevant to the demands of their jobs -- including, for example, calling parents, presenting lessons, formulating questions, translating letters, and correcting students' writing. One section however, the written reaction to a professional reading, was not perceived to be part of a bilingual teachers' day according to the teaching experts.

Construct validity. In order to answer this question (Do the oral and written parts of the test provide different information?) an exploratory factor analytic procedure (Principal Factor Analysis and Promax rotation) was selected. All of the variables (that is, the test sections) were entered into the analysis in their most unchanged form (not summed for scale scores related to vocabulary, appropriateness, and so on).

The results from the procedure are found in Table 6. Of interest for this study are the factor loadings at or above .30 (Biber, 1988). Factor 1, made up only of the written sections, supports the notion that the written part of the test yields information that is different from the other factors.

Factor 2 and Factor 3 however, both consist of sections relating to the oral part of the test. This suggests that the oral part of the test is different from the written part, but that some of the oral sections (Section 1, Section 2, Section 9) may be different from the remaining oral sections (Section 3, Section 4, and Section 5). Closer examination of the sections revealed that the major difference between these two oral groups is that the first requires the examinees to use language under relatively realistic time constraints. This pressure to produce responses orally results in language samples that can be characterized as spontaneous. That is, examinees have virtually no planning time for their responses.

The second set of oral sections which separated out from the first oral factor and the written factor involve those test tasks that require the examinee draw on knowledge, especially lexical and syntactic knowledge, which has to do with teaching Spanish. These tasks allow the examinees ample time to plan carefully how to respond. This situation is quite different than the oral tasks represented by the first oral factor above. Indeed, it could be argued that this factor is more like the written factor than it is like the "spontaneous" oral factor, at least where planning time is involved.

Table 6

Rotated Factor Pattern for ACTSPE

Exam section/subsection	Factor 1	Factor 2	Factor 3
Sec.1 Appropriateness average	-0.08	0.73	0.14
Sec. 1 Grammar average	0.05	0.81	-0.00
Sec. 1 Vocabulary average	0.08	0.85	-0/06
Sec. 1 Naturalness average	-0.09	0.56	0.32
Sec. 2 Average	0.12	0.53	0.13
Sec. 9 Average	0.08	0.37	-0.04
Sec. 3 Comprehension average	-0.07	0.03	0.60
Sec. 3 Fluency average	0.05	-0.06	0.81
Sec. 3 Grammar average	0.25	0.22	0.48
Sec. 4 Average	0.10	0.16	0.57
Sec. 5 Average	0.12	0.16	0.46
Sec. 6 Grammar average	0.67	0.16	0.01
Sec. 6 Spelling average	0.63	-0.10	0.07
Sec. 6 Vocabulary average	0.79	0.07	-0.10
Sec. 6 Style average	0.71	0.23	-0.14
Sec. 7 Comprehension average	0.45	-0.08	0.29
Sec. 7 Writing average	0.76	-0.05	0.10
Sec. 7 Style average	0.69	0.06	0.07
Sec. 8 Average	0.44	0.05	0.18

Discussion and Implications

The results of this study indicate that, overall, the ACTSPE is a reliable and valid measure of teacher language proficiency. The analyses addressing the reliability of the ACTSPE indicated that the oral and written parts are consistently measuring oral and written ability, respectively. In addition, the raters at NAU are scoring the ACTSPE consistently overall, although some rater training is suggested, particularly for pair 3 and 4.

The analyses addressing aspects of validity all provided evidence that the test is measuring what it is intended to measure. Not only was the test perceived to be appropriate in difficulty and as measuring the linguistic abilities needed for teaching, but the expert teachers who were interviewed felt that the majority of the content was representative of what they had to do in the bilingual classroom. Additional evidence that each main part of the test is providing the evaluator with different information was found using the factor analytic procedure.

Exam content

While the study of the ACTSPE, taken as a whole, suggests that the test is reliable and valid, the results from two of the research questions merit further discussion. First, for the research question which addressed evidence for content validity, the experienced bilingual teachers indicated that Section 7, Written summary and reaction to professional reading, is not representative of a bilingual teachers' duties. The task's purpose however, according to the test developers (Norfleet, 1994) was to have the examinee demonstrate a sophisticated level of reading and writing by a two part item. That is, after reading the professional article, examinees must first summarize the article and then write a reaction to it. Clearly the teachers in this study do not do this regularly. Most of their reference materials are in English and they seldom summarize or formally react to them (except, perhaps, in teacher training courses). A tension therefore exists between the teachers and their reported duties and the developers of the test:

teachers indicated that the professional reading and summary was not representative of their "duties" while test developers felt that the ability to perform such tasks was essential.

Test tasks: general or specific?

Related to these observations about exam content is the issue of testing language ability or testing teachers' language ability with respect to the tasks they must potentially deal with in the classroom. As mentioned above, the ACTSPE has been purposefully designed to include content that *should* be a part of the prospective bilingual teachers' repertoire. Teacher training programs however, don't always include this type of training. In a survey of 22 prospective bilingual teachers, done by Gonzales (1993), it was determined that one-third to one-half of respondents had not had courses preparing them for the bilingual classroom. They lacked foundation courses in bilingual education, methodology courses in bilingual education, and education courses in Spanish. In fact, many of the respondents said that they never a teacher that spoke Spanish during their training. As a result, students involved in the training described above would be at a disadvantage on a test like the ACTSPE in which the test tasks are closely tied to classroom tasks.

Further study is needed to determine whether it is preferable to test teachers by using more general measures of language proficiency or more specific measures of teachers' ability to use language in classroom "scenarios."

Oral factors

Another set of results worthy of further discussion is the outcome of the factor analysis. In particular, the separation of the oral part of the test into two factors indicates that the oral part of the test is measuring something more than general oral ability. The first oral factor consists of the test sections that require the examinees to produce and/or comprehend speech within "real-life" time constraints. This use of language might initially suggest that the first oral factor taps aspects of sociolinguistic competence (Bachman, 1990). The scoring rubric, however, indicates that this view is only partially true. While many of the sections making up the first oral factor are

scored for appropriacy, they are also scored for grammar and vocabulary. These tasks, then, require an integration of competencies, modeling the way that language is used in real life.

The remaining oral sections, making up the second oral factor, may consist of tasks dealing with the planned production of language. Especially salient are the sections which require the examinee to produce appropriate lexical items. While this aspect may lead one to conclude that the sections are tapping grammatical competence (Bachman, 1990), closer scrutiny of the sections again reveals that this interpretation is too simplistic. Further examination of the sections again shows that there is much more required to perform the tasks than solely grammatical competence.

As a result, the most useful interpretation of the two oral factors has to do with the way in which language is processed and/or produced in the oral sections. In the sections which make up the first oral factor, the examinee had to perform those tasks on-line, relatively spontaneously, under time constraints. In general, these sections required the examinee to produce and/or comprehend at a conversational rate that allows neither replaying nor retaping.

On the other hand, the second oral factor, comprised of the remaining oral sections, involved tasks where the examinees had ample time to think about their responses and there was no "on-line" processing involved. They could both replay their tapes and retape their responses as these sections allowed examinees to proceed at their own individual rates.

This differentiation of oral tasks needs further study in order to rule out what may be primarily a method effect. Some research indicates that oral language production which is planned is quite unlike that which is spontaneous (see for example the transactional/interactional distinction made by Brown and Yule, 1983, and the involved versus informational production dimension posed by Biber, 1988). These differences might also be compared to the basic interpersonal communicative skills/cognitive-academic language proficiency (BICS/CALP) distinction described by Cummins (1984). Indeed, Collier (1985) points out that it may be important to measure both BICS and CALP, "including CALP in the content areas the teacher will be teaching" (p. 9).

However, is could also be argued that the BICS/CALP (as well as transactional/interactional language use and Biber's involved versus informational production dimension) may in fact represent continuums of language use that teachers draw on in different ways. Much of teaching is individual style: the ability to relay important facts to students in an informal way may be just as effective as teaching content more formally. In other words, both uses of language are manifested, but to different degrees according to individual teaching style (which most likely be affected by the makeup of the class and individual learning styles). Therefore, *both* BICS (as well as interactional or involved language use) *and* CALP (as well as transactional and informational language use) are part of a teacher's proficiency; the degree to which they are drawn on reflects the teacher's preference, style, and the task/context.

The notion that different language situations draw on different language abilities is borne out by the results of this study; how separate, or individual, those abilities are must be researched further. Nevertheless, authors of teacher training texts such as Richards (1990) and Nunan (1991), point out that learners of a second language need instruction and practice in both types of language use, whether they be labeled as transactional/interaction or informational/involved. This application should not stop at teaching language, however. It is also informative for teacher preparation and teacher assessment.

Teacher training and assessment

Given that this test was developed from a needs analysis of bilingual teachers, it follows that the tasks reflect those faced by these teachers. The research question which asked bilingual teachers to comment on the test content supports this. Interestingly enough, however, these test tasks, when subjected to the factor analysis, separated out in a way that indicates that the test is capturing information about an examinee's ability to write, ability to speak spontaneously, and ability to execute planned speech. Based on the results of this study, more research is needed; studies which investigate bilingual teachers' use of language in the classroom as well as analyses of speech and writing samples will help shed light on these proposed different types of language use.

In more general terms, the results of this study are informative for the larger picture of teacher testing outlined in the beginning of this paper. Numerous states offer bilingual certification and/or endorsement, but just over half require assessment of teachers' language abilities, some using single skill measures while others use multiple tests. This study suggests that, as a minimum, the competencies needed by a bilingual teacher in the classroom, including spontaneous and planned oral ability as well as written ability, should be assessed; measures tapping less are not enough.

Furthermore, for prospective teachers, appropriate language training (including using language as the medium to teach content as well as communicating with parents and the community) should be part of the teacher preparation programs. Future teachers need the opportunity to learn and put into practice the language abilities that they will draw on, both in the classroom as well as out of the classroom.

Conclusion

The importance of the ACTSPE as an endorsement test as well as other teacher assessment procedures cannot be underestimated. Decisions based on these tests have far reaching implications: they affect students' exposure to appropriate role models where language is concerned as well as teachers' careers, salaries, and job security. And, although the need for bilingual teachers is very great, it is unfair to the teachers, parents, and students alike, *not* to require potential teachers to meet a certain minimum standard of language ability. The standards set not only affect the quality of Spanish used in the classroom, but also the attitude toward bilingual education in the community.

Bilingual teacher assessment currently takes a variety of forms. And, although it is often easier to use a single skill measure, or disregard testing altogether, it must kept in mind that the trend toward bilingual and multilingual classes is not going to change. The increasing numbers of

LEP students require that teachers be appropriately prepared to enter the classroom, and that preparation must include competence in the second language.

References

American Council on the Teaching of Foreign Languages. (1986). ACTFL proficiency guidelines. New York: American Council on the Teaching of Foreign Languages.

Arizona Department of Education. (1992). Bilingual programs and English as a second language programs. Annual report Phoenix, AZ: Arizona Department of Education.

Bachman, L. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

Barkin-Reigelhaupt, F. (1985). Testing bilingual language proficiency: An applied approach. In Elías-Olivares, L., Leone, E., Cisneros, R. & J. Gutiérrez (Eds.) Spanish language use and public life in the United States (165-180). NY: Mouton Publishers.

Biber, D. (1988). Variation across speech and writing. NY: Cambridge.

Brown, G. & Yule, G. (1983). Discourse analysis. NY: Cambridge.

Collier, V. (1985). University models for ESL and bilingual teacher training, Arlington, VA (ERIC Document Reproduction Service No. 273 154).

Cooperative Personnel Services (1992). Bilingual Certificate of Competence Examination (Spanish). Examination Bulletin. Sacramento, CA.

Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In Rivera, C. (Ed.) Language proficiency and academic achievement (2-19). England: Multilingual Matters Ltd.

Esquivel, R., Ovard, J., Wooten, H., & Wilkison, C. (1990). Staffing the multilingually impacted schools of the 1990s. Washington, D.C.: U.S. State Department of Education, Office of Bilingual Education and Minority Languages Affairs.

Gonzalez, F. (1993). Creating education that works: Building bilingual teacher competencies. Intercultural Development Research Association, San Antonio, TX (ERIC Document Reproduction Service No. 366 483).

McFerren, M., Valadez, C.M., Crandall, J., Palomo, R.S. & Patiño-Gregoire, C. (1988). Certification of language educators in the United States. University of California, Los Angeles. (ERIC Document Reproduction Service No. 291 244).

National Clearinghouse for Bilingual Education. (1991). [State description of Bilingual Education certification and exams required]. Unpublished raw data.

Norfleet, L. (1994). Reliability and validity of the Arizona Classroom Teacher Spanish Proficiency Exam. Doctoral dissertation, Flagstaff, AZ: Northern Arizona University.

Nunan, D. (1991). Language teaching methodology. NY: Prentice Hall.

Richards, J.C. (1990). The language teaching matrix. NY: Cambridge.

SAS Institute Inc. (1985). Cary, NC: SAS Institute Inc.

Waggoner, D. (1992). Numbers and needs: Ethnic and linguistic minorities in the United States, Washington, DC (ERIC Document Reproduction Service No. 359 285).

Wesche, M. (1992). Performance testing for work-related second language assessment. In E. Shohamy & A. Walton (Eds.) Language assessment for feedback: Testing and other strategies (pp. 103-122). Dubuque,Iowa: Kendall/Hunt.

Appendix A

Arizona Classroom Teacher Spanish Proficiency Exam -- Sample questionnaire items

1. What is your first language? _____

2. Where did you learn Spanish? _____

3. How would you rate the overall level of difficulty of this examination?

   a. too difficult _____
   b. appropriate _____
   c. too easy _____

4. How useful do you think this examination is for diagnosing the linguistic abilities of the teachee in a bilingual setting?

   a. very useful _____
   b. moderately _____
   c. not at all useful _____

5. My ability to speak Spanish with a native speaker is:

   Poor__ Marginal____ Fair____ Good____ Excellent____

6. My ability to comprehend Spanish spoken by a native speaker is:

   Poor____ Marginal____ Fair____ Good____ Excellent____

7. My ability to comprehend written (formal) Spanish is:

   Poor____ Marginal____ Fair____ Good____ Excellent____

8. My ability to write in Spanish is:

   Poor____ Marginal____ Fair____ Good____ Excellent____

Appendix B

Teacher expert questionnaire

Directions: Please answer the following question for each number by circling Y (yes) or N (no).

Is the task described in each section something that a bilingual teacher would have to perform?

1. Telephone call to parent Y N

2. Teacher reading aloud to students Y N

3. Teacher presenting a math lesson Y N

4. Teacher formulating questions based on story Y N

5. Use vocabulary such as:

   a. math terms, punctuation Y N

   b. dates Y N

   c. commands, expressions Y N

   d. terms for relatives Y N

6. Teacher translating a letter home Y N

7. Summarize and react to a professional article Y N

8. Correct students' writing Y N

9. Listen to children discuss story or play Y N